

Contrast Maintenance in Language and the Innateness Debate.
Andrew Wedel, March 2005

Two articles have recently appeared ('Linguistics Research Moving in New Direction', February 21, 2005. Harrison, J., UANews, Tucson; 'Is this the last word in linguistics?' March 2, 2005. The Independent, UK) describing some of my work in terms of a conflict between 'innatist' and 'emergentist' hypotheses concerning the origin of language structure. Summarizing work to create a news story inevitably involves simplification, and to respond to some of those that I think are more significant, I'd like to re-summarize my work in this context, and explain how I think it contributes to this larger theoretical conversation.

Although human languages show a remarkable diversity in their sound and sentence patterns, linguists also identify many features that languages usually share. As an example of a universal feature, in the domain of phonology (my specialization), we know that all languages, spoken or signed, construct words out of re-usable parts, rather than creating a unique sound or gesture for every word. As an example of a common, but not universal feature, almost all languages that use voiced stop sounds like [b], [d] and [g], also use the corresponding voiceless stop sounds, in this case [p], [t], and [k].

How do these commonalities arise? One extreme hypothesis would be that all universals and tendencies are 'hard-wired' into our genes – that is, that these properties of language are altogether innate. At the other extreme, we could propose that the neural hardware we use for languages doesn't constrain languages in these ways whatsoever, and that these commonalities emerge through the influence of extra-genetic processes and other facts-about-the-world. These two extreme hypotheses can be thought of as defining abstract poles on a continuum of possibilities from 'completely innate' to 'completely emergent'.

However, as has been cogently argued many times before (e.g., Scholz, B. C. 2002. Innateness. *Nature* 415: 739-739.), neither of these two extreme hypotheses are really useful in themselves because the no human behavior is ever fully one or the other. Every behavior uses our bodies in some way, and so is genetically constrained at some level, while every phenotype must emerge from a more or less indirect mapping from genes, and as such is always subject to extra-genetic modification. As an example outside of language, consider the way I hold a fork. It seems highly unlikely that the way I hold a fork is genetically encoded in any specific way, because forks haven't been around long enough for any difference in the reproductive success of people who hold forks in this or that way to have influenced the gene pool. On the other hand, no human is born with completely free choice about how to hold a fork either – the genetically specified form of our hands make some solutions likely, others impossible or highly unlikely. As such, a good hypothesis would be that the specific way I hold a fork is the result of culture, acting within a highly constrained set of possibilities defined by my genes.

How does this fit into language? The generative model of language, supported by Noam Chomsky among many others, holds that the many strong patterns of similarity across languages, despite their awesome complexity, can only be the result of a high degree of genetic pre-specification, in the form of a uniquely human 'universal grammar' deriving from an innate language organ in our brains. However, a great deal of research has also shown that many kinds of reproducible, complex structures in the physical and

biological worlds can come about without extensive pre-specification. These complex structures are created through 'self-organization', which occurs when local feedback loops cycling through a simple system over time result in the slow accumulation of longer-range structure. Famous examples in the physical world include the hexagonal structure of honeycombs, ripples on a sandy creek bed, and the characteristic shape of thunderstorms. Many patterns within culture have also been proposed to emerge through self-organization operating over individuals and groups, including patterns of buying and selling within the stock market, and persistent racial segregation in neighborhoods (for two good reviews, see Rauch, J. 2002. Seeing around corners. *Atlantic Monthly*, April 2002; Kaufmann, S. A. 1995. *At home in the universe: The search for the laws of self-organization and complexity*. Oxford: Oxford University Press).

Many linguists have noted that language is a system in which the general prerequisites for self-organization induced structure are richly met: there are lots of little interacting elements at many levels, and there are many possible feedback loops driven by repetition and copying within generations of language speakers, and within language transmission across generations. As a consequence, it makes sense to look for ways in which some of the reproducible, complex patterns we see in language might arise via this general mechanism. To ask this question is not to deny that there are physical and genetic constraints on language patterns. Rather, asking this question acknowledges the fact that in the mapping from genes to phenotype, there are many different points at which phenotypic complexity can arise.

Here is an example from my current work that illustrates the potential of this research program. A long-standing problem in phonology involves the long-term maintenance of sound contrast within a language. Given that we tend to progressively cut corners in speech, and that sounds do often merge together over time, how do languages avoid devolving to complete homophony? Why do we not eventually all just say 'bababa' to each other? In a functional sense, the answer is obvious: if all our words sounded the same then we wouldn't be able to say anything. But the actual mechanism for keeping words and sounds apart over the course of language change isn't so obvious.

Theoretical models over the last century that have grappled with this problem have all, in essence, ended up postulating some innate 'watchdog' function of the grammar which overtly monitors sound contrast. Sometimes, this watchdog intervenes when two kinds of sounds get too similar and moves them back apart – and sometimes it doesn't. For example, on the American East Coast, the vowels that appear in the words 'cot' and 'caught' are different, while in many dialects further west, these two vowels have merged, such that 'cot' and 'caught' now sound the same. In this particular case, the watchdog in the American West seems not to have bothered to keep these two originally contrastive vowels from merging. If the contrast maintaining mechanism is a highly pre-specified part of 'universal grammar', how can we understand the fact that in this case these vowels were permitted to merge, when in many other cases of language change, vowels are kept perceptually distinct?

In recent work, I've articulated an alternative model, in which the maintenance of sound contrast over time is indirectly driven by the functional role of distinct sets of sounds in conveying distinct meanings. This model is based on the recognition of a subtle statistical effect in the act of mapping a perceived word to a meaning category that favors contrastive word variants over less contrastive word variants. Because of this slight

advantage, over many, many cycles of speech, contrastive word forms end up having a greater influence on the evolution of the language than less contrastive word forms, indirectly resulting in a tendency to preserve contrast between most sounds over the long-term.

Like all self-organizational phenomena, this phenomenon depends on a feedback loop. Here's how this feedback loop operates. There is a great deal of experimental evidence that our sound-meaning categories contain a rich level of phonetic detail. Further, experiments show that our pronunciation of a word is influenced by the pronunciations of that word that we've recently heard. The influence of experienced detail on pronunciation can be modeled by making the simplifying assumption that we literally remember the fine details of every word we hear, and that any word that we've heard and remembered can serve as a model for our future pronunciation of that word. Now, if we remember phonetic details of words that are spoken to us, and these phonetic details can emerge in our own pronunciation, a positive feedback loop is set up in which a common pronunciation will tend to spread in a community at the expense of less common pronunciations.

Imagine now that we have two words, A and B, that sound nearly, but not quite the same. Spoken variants of A that sound *less* like B will still be recognized as A by a hearer, but variants of A that sound *more* like B may well be mistaken as B. The same goes for variants of B: those that sound *less* like A will still be understood as B, but those that sound *more* like A may well be understood as A rather than B. Given positive feedback between perception and future pronunciation, random variation in pronunciations will therefore tend to push the 'usual' pronunciation of A and B further away from one another with time, even though there is no mechanism in the model that directly acts to protect the difference in pronunciation between A and B.

To illustrate this effect, I set up a simulation with two speakers, who have little toy lexicons comprising a set number of sound-meaning categories. Each category contains examples of 'words' that have been heard before; each word is built out of sequences of sounds located in a small 2-dimensional phonetic space – we could think of these sounds as vowels. They take turns speaking all the words in their lexicons to each other over many, many rounds of conversation, and continually try to categorize what they've heard on the basis of what they already have stored in each of their lexical categories. Each time they hear a word, they store it in the category that matches best, and each time they speak, they make some small mistakes in pronunciation, with the result that the sounds in their words slowly, randomly drift around with time. However, the differences in confusability of different variants, as discussed above, conspire over the long run to keep the speakers' words distinct. Every once in a while, two words do become homophonous, but much less often than one would expect by chance. As a result, the lexicons of the two speakers retain contrast over time, even though their 'sound-systems' slowly change. This looks qualitatively like the pattern we find in human languages.

Now what have I shown here? I have not shown that contrast maintenance is not 'innate'. I have just potentially pushed back the specificity of that innateness, from a highly specified watchdog of contrast in the grammar, to a more general property of our innate categorization behavior. However, even while this mechanism remains 'innate' at some level, it is also emergent: contrast maintenance only develops through many rounds

of interactive communication between multiple speakers. As a result, this mechanism cannot be defined solely in terms of an innate, physical structure residing within a single person's brain – it is in important ways dependent on community, history, and chance, and therefore can be thought of as being distributed both over speakers, and time. Just as the way I hold my fork is likely to be influenced by culture as well as genes, this model proposes that contrast maintenance in language is the result of complex and long-term interactions within a linguistic community, traversing pathways defined by our innate categorization mechanisms.